Computers & Education 126 (2018) 129–142

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Virtual learning environment engagement and learning outcomes at a 'bricks-and-mortar' university

Chris A. Boulton*, Carmel Kent, Hywel T.P. Williams

College of Engineering, Mathematics and Physical Sciences, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4QE, UK

ARTICLE INFO

Keywords:
Interactive learning environments
Human-computer interface
Learning communities
Post-secondary education

ABSTRACT

In this study, we analyse the relationship between engagement in a virtual learning environment (VLE) and module grades at a 'bricks-and-mortar' university in the United Kingdom. We measure VLE activity for students enrolled in 38 different credit-bearing modules, each of which are compulsory components of six degree programmes. Overall we find that high VLE activity is associated with high grades, but low activity does not necessarily imply low grades. Analysis of individual modules shows a wide range of relationships between the two quantities. Grouping module-level relationships by programme suggests that science-based subjects have a higher dependency on VLE activity. Considering learning design (LD), we find that VLE usage is more important in modules that adopt an instruction-based learning style. We also test the predictive power of VLE usage in determining grades, again finding variation between degree programmes and potential for predicting a student's final grade weeks in advance of assessment. Our findings suggest that student engagement with learning at a bricks-and-mortar university is in general hard to determine by VLE usage alone, due to the predominance of other "offline" learning activities, but that VLE usage can nonetheless help to predict performance for some disciplines.

1. Introduction

The relationship between engagement and learning outcomes has been investigated in both online learning (Adeyinka & Abdulmumin, 2011; Andresen, 2009) and in 'bricks-and-mortar' (physically situated, face-to-face, predominantly offline) educational institution settings (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Carini, Kuh, & Klein, 2006). One reason for the general lack of clear evidence about this relationship (Picciano, 2002) is that the concept of student engagement has been differently defined (Gunuc, 2014) and operationalized (Akyol & Garrison, 2011) in different contexts. One definition for "student engagement" has come to refer to the level of involvement students appear to have within their classes and their institutions in the context of learning (Axelson & Flick, 2010). Student engagement in higher education has been operationalized in various ways. Commonly used methods for determining engagement include self-reporting by standard surveys, such as the National Survey of Student Engagement (NSSE) (Carini et al., 2006) and the College Student Experiences Questionnaire (CSEQ) (Axelson & Flick, 2010), and digital traces such as VLE usage logs (Agudo-Peregrina et al., 2014). In this paper, we argue that student engagement at traditional bricks-and-mortar higher education institutions is a complex and multi-dimensional construct (Casuso-Holgado et al., 2013) entailing the measurement of students' interactions with various types of resources and agents (such as systems, people and devices) associated with the individual learning experience (Shavelson & Huang, 2003). Thus in order to evaluate the relationship between engagement and outcome in a bricks-and-mortar institution, studies cannot rely on a single or a reduced number of proxies for

* Corresponding author.
E-mail addresses: c.a.boulton@exeter.ac.uk (C.A. Boulton), kent.carmel@gmail.com (C. Kent), h.t.p.williams@exeter.ac.uk (H.T.P. Williams).

https://doi.org/10.1016/j.compedu.2018.06.031
Received 4 September 2017; Received in revised form 27 June 2018; Accepted 30 June 2018
Available online 02 July 2018

engagement. Even in domains other than learning, it is well established that engagement should not be measured through one fixed set of metrics, but rather with context-dependent models (Lehmann, Lalmas, Yom-Tov, & Dupret, 2012). Student engagement is highly sensitive to the learning context, specifically to the chosen LD (Rienties & Toetenel, 2016) and assessment method (with LD being highly co-dependent) (Nguyen, Rienties, Toetenel, Ferguson, & Whitelock, 2017). With this high co-depenency, this relation cannot be studied or reported without taking into account the variability in LD and disciplinary contexts.

In this paper we analyse relationships between VLE usage and grades for a range of programmes at a traditional bricks-and-mortar university in the UK, also considering the effects of LD and disciplinary context. Our findings suggest that student engagement is hard to determine by VLE usage alone (in contrast to stronger relationships found in other contexts such as online and distance learning), due to the predominance of other "offline" learning activities. In addition, we show that the predictive power of VLE usage for learning outcomes varies between the various programmes, which might be explained by different LDs, as well as how VLEs are used within each degree programme. Finally, we fit a predictive model on VLE usage in each degree programme to determine how well grades can be predicted on VLE usage alone.

We hope that this study provides insights into the complexities of measuring learning behaviours and determining engagement at a bricks-and-mortar university, as well as the substantive variation in learning activities between degree programmes. Using similar models to ours, we believe that decision makers might be able to identify potential 'at-risk' students weeks in advance of assessments, so that additional support can be given if needed.

2. Background

We introduce our study by reviewing the distinctive nature of student engagement in bricks-and-mortar learning settings when compared to online-only and blended offline/online settings. Bricks-and-mortar settings impose different challenges for measurement of engagement and its relation with learning outcomes. We then continue to briefly review the context-dependent nature of engagement, as it relates to pedagogical designs of learning activities.

2.1. Measuring the relationship between engagement and outcome at a bricks-and-mortar university

Engagement and participation in online learning environments have been shown to be highly correlated with various types of learning outcomes (Adeyinka & Abdulmumin, 2011; Asterhan & Hever, 2015; Cerezo, Sánchez-Santillán, Paule-Ruiz, & Núñez, 2016; Kuzilek, Hlosta, Herrmannova, Zdrahal, & Wolff, 2015; Song & McNary, 2011; Zhu, 2006). In purely online learning settings, engagement is usually defined in terms of student interactions with a single system, typically a VLE. However, student engagement at bricks-and-mortar institutions, where most teaching is delivered face-to-face, and in which engagement has historically been teacher-centric, is much less clearly defined. Online learning is not analogous to face-to-face learning and each requires different conceptualization and operationalization frameworks (DeBoer, Ho, Stump, & Breslow, 2014; McConnell, 2000). Students are shown to engage differently when learning in an online learning environment as opposed to a bricks-and-mortar environment, resulting in different learning outcomes and a different set of variables to predict outcome (Agudo-Peregrina et al., 2014; Harris & Nikitenko, 2014). The emergence of online learning fostered a shift towards the decentralization of the learning process, which resulted in a stronger emphasis on course structure and content, and higher volume of interactions with digital systems, while leaving in-class learning to be more complex and difficult to systematically quantify by nature (Agudo-Peregrina et al., 2014). Studies using self-reporting instruments (such as surveys) to measure engagement in a bricks-and-mortar environment reported on overall positive but rather weak relations and limited ability to predict outcome across contexts (Carini et al., 2006; Casuso-Holgado et al., 2013; Gunuc, 2014). Students' VLE logs have been used as an alternative proxy for engagement in order to overcome the natural biases of self-reporting, showing again to be limited and less predictive in face-to-face settings then in online settings (Agudo-Peregrina et al., 2014).

2.2. Engagement is context-dependent

Higher education is highly varied in context, including the physical environment, digital systems and resources, the indicators used for assessment, and the pedagogical approaches used (Harris & Nikitenko, 2014). An LD, chosen and executed by the instructors, represents the collection of learning activities and objects that are used during the execution of these activities (e.g. books, articles, software programmes, pictures). It can also refer to services (e.g. forums, chats, wikis) used during these activities (Koper, 2006) and might also include the instruments chosen to assess outcomes (Millar, 2013). The diversity of constituents creates a very high diversity of possible LDs.

Although the importance of LD for conceptualising and advancing the quantitative analysis of learning has been recently recognized, very few studies to date have investigated a substantial number of LDs with VLE proxies and learning outcome (Rienties & Toetenel, 2016). In the context of online and blended learning, LD, disciplinary differences and assessment instrumentations were able to explain a high portion of the variability in VLE engagement (Nguyen, Rienties, & Toetenel, 2017; Nguyen, Rienties, & Toetenel, 2017; Rienties & Toetenel, 2016). LD elements such as assessment and course length were also shown to influence engagement in Massive Open Online Courses (MOOCs) (Ferguson et al., 2015). In the context of bricks-and-mortar learning, LD was shown to influence students' approaches to learning (Wilson & Fowler, 2005) and engagement (Ahlfeldt, Mehta, & Sellnow, 2005). It was also found that the level of engagement depends on the college or subject (Ahlfeldt et al., 2005), which might also be explained by different LDs and instructional methods.

Table 1
Programmes selected for our study from which 1st year compulsory modules are taken.

College	Programme	Number of 1st year compulsory modules	Number of unique students taking at least one module
College of Life and Environmental Sciences	BSc Biological Sciences	8	252
College of Humanities	BA English	4	353
College of Engineering, Mathematics and Physical Sciences	BSc Mathematics	8	352
College of Social Sciences and International Studies	BA Politics	8	523
Business School	BSc Economics	6	532
University of Exeter Medical School	BSc Medical Sciences	4	158

3. Dataset and methods

3.1. Selection of data for analysis

Our main data is collected from 2015/2016 academic year at the University of Exeter (UoE), an established bricks-and-mortar higher education institution in the UK. The UoE is organised academically as six Colleges, each of which offers a number of undergraduate and postgraduate degree programmes in several disciplines. Here we consider undergraduate programmes, which consist of 360 credits earned through successful completion of taught modules at different levels (typically 120 credits per year for three years of study). Some modules are compulsory for each programme, while others are optional and selected according to student preference. To ensure a diverse disciplinary spread of degree programmes for our analysis, we chose a single programme from each college (Table 1). For each programme, we chose all the compulsory modules for the first year of study; this selection served to reduce the variation introduced by optional modules and also ensured a large student cohort for each module studied. We extracted data from all students who took each module in the 2015/16 academic year. We include all students who took the module, regardless of if they were taking the programme for which the module is compulsory (this implies that some students may be taking a module as an optional component of a different programme, but this is rare in the first year). We note that there are 2025 unique students in our study despite the sum of the final column in Table 1 equalling 2170 as some students take modules from more than one programme (e.g. there is some overlap in the BSc Medical Science and BSc Biological Sciences programmes).

Of the 38 modules analysed here, 2 BA English modules, 2 BSc Economic modules, and all 4 BSc Medical Sciences modules were worth 30 credits (1/4 of a student's total credits for the year), with the rest being worth 15 credits (1/8 of the total credits for a year). The number of students taking individual modules ranged from 151 for the module ‘Plants’ in BSc Biological Sciences and 349 for the modules ‘Beginnings’ in BA English.

As a test dataset for our predictive model, we also use data from the same modules described above but for the 2016/17 academic year (see Section 3.5). To avoid confusion with our main dataset from the 2015/16 academic year, we do not list the numbers of students from the 2016/17 dataset here but can confirm the cohort sizes are very similar.

3.2. VLE usage data

The VLE at the UoE has a hub for each module where lecture slides, worksheets and extra reading material is uploaded by the lecturer. There is also a forum space for discussion of module content and limited functionality for assessments (e.g. class tests, quizzes). The VLE is run on the MOODLE software platform (https://moodle.org), which provides a timestamped log of every student interaction with the system, whether this be viewing course materials, interacting with the forum, or any other activity. Each interaction is associated with a particular module.

Using the VLE log data we create a time series for each student's activity associated with each module in our sample by summing the total number of minutes per day that they are active on the module in question (Tempelaar, Rienties, & Giesbers, 2015). We say a student is active in a module if their activity sequence shows multiple interactions with that module on the VLE within a 5 min period. If there is no activity after 5 min, we estimate they were active on the final page for 2 min. That is, their active period begins at the first activity timestamp and terminates 2 min after the last activity timestamp prior to an inactive period of at least 5 min. If they become inactive in a module by becoming active in another module, we count the time up to the change as being active in the first module. Using these rules, we are able to estimate the number of minutes each student was active on each day for a given module.

Modules are typically taught across a single 11-week term (15 credit modules) or across two 11-week terms (30 credit modules) in a single academic year (in some cases 30 credit modules can be taught in a single term). Although exams are generally taken a few months after the end of the taught section of a module, we focus on student engagement during the term(s) in which the module is taught. After creating time series of student VLE activity per day for each module, we are able to calculate the mean time per day that a student engages with a module across the one or two terms in which the module was taught.

3.3. Module grades data

Students are awarded a percentage grade for each module they complete. This grade is calculated from a weighted average of

Table 2
Learning design (LD) variables extracted for each module (n = 38 modules).

Learning design variable	Mean (n=38)	Std. Dev.	BSc Bio. Sci. (Mean)	BA English (Mean)	BSc Math. (Mean)	BA Politics (Mean)	BSc Econ. (Mean)	BSc Med. Sci. (Mean)
Number of students	223.11	52.23	197.50	243.25	240.63	211.75	279.50	157.25
Number of scheduled teaching activities hours (per credit)	2.38	0.64	2.76	1.79	3.00	1.78	1.89	2.87
Total number of scheduled lectures hours (per credit)	1.36	0.54	1.37	0.57	1.91	1.11	1.54	1.22
Number of scheduled seminars hours (per credit) (part of the total number of scheduled lectures hours)	0.27	0.48	0.05	1.19	0.21	0.08	0.07	0.57
Total number of guided independent study hours (per credit)	7.62	0.64	7.24	8.21	7.00	8.22	8.11	7.12
Number of VLE tasks hours (per credit) (part of total number of guided independent study hours)	0.12	0.44	0.16	0	0	0.33	0	0.14
Weight of course work in summative assessment of credit	30.74	25.69	27.87	60.00	21.25	46.25	5.00	33.75
Weight of written exams in summative assessment of credit	64.97	27.93	62.37	30.00	78.75	53.75	93.33	57.50
Weight of practical exams in summative assessment of credit	4.29	8.45	9.75	0	0	0	1.43	8.75
Weight of VLE tasks in summative assessment of credit (detailed division)	2.63	7.95	1.25	10.00	0	0	5.00	5.00
Weight of practical report in summative assessment of credit (detailed division)	1.97	6.93	6.87	0	0	0	0	5.00
Weight of scientific essay in summative assessment of credit (detailed division)	18.84	26.58	10.37	50.00	0	50.37	0	7.5
Weight of participation in summative assessment of credit (detailed division)	1.05	3.11	0	7.50	0	0	1.67	0

multiple module assessments, including (e.g.) essays, multiple choice tests, coursework, or a final exam. The number, type and weighting of assessments varies between modules.

3.4. Learning design metadata

Different disciplines commonly follow different LDs, which employ different combinations of assessment types and time spent on “contact” learning (e.g. lectures, classes) versus “self-study” learning (e.g. reading). For example, modules from “Arts” programmes (such as BA English) are more likely to have more essays and fewer exams, whereas “Science” programmes (such as BSc Mathematics, BSc Biological Sciences) often have a mixture of class tests and practical work, alongside a final exam.

At the UoE, each module designer (or instructor) is required to publish a basic list of structural features, some of which are standardized across different departments, which can be used as proxy variables to characterise LD. We extracted a range of LD variables from the UoE module catalogue as detailed in Table 2 below. The time spent in each module is typically divided into hours spent on “scheduled teaching activities” and hours spent on “guided independent study”. The typical expectation at the UoE is that 1 credit corresponds to around 10 h of work by the student, so a 15-credit module will typically correspond to around 150 h of study. Student performance on each module is measured by a number of “summative” assessments which contribute to an overall percentage mark as described above. The balance of different assessment types is at the discretion of the module designer.

3.5. Model on VLE usage to predict grades

We fit ordinal logistic regression (OLR) models to our 2015/16 dataset to predict a student's module grade based on their mean daily VLE usage. The output of these OLR models give a probability of the results of a student's module being within one of a number of ordered categories. We chose our categories as ‘Fail’ (< 40%), ‘3rd/2:1’ (> 40%, < 60%), ‘2:1’ (> 60%, < 70%), and ‘1st’ (> 70%), as they are important grade boundaries at UK universities. Once we have fitted the OLR models to our dataset (one for each degree programme), we use the regression coefficients to explore how the model predicts changes to the probability of being in each grade boundary as the number of VLE minutes per day is increased.

To determine how well our models predict grades, we use the equivalent data for the 2016/17 academic year as a test set and determine how well our models (trained on the 2015/16 data) can predict grades for students in the 2016/17 cohort. This test is intended to give some indication of how well this approach could be operationalized, e.g. to provide information to tutors about students that may be at-risk of failing a module. In that scenario, we assume that predictions must be made for the current year using models trained on data from previous years. Goodness of fit of these models is measured with a modified residual sum of squares (RSS) score, detailed below. RSS scores measure the discrepancy between data and an estimation model, and are often minimised as an optimisation criterion.

For each grade boundary, we discretise the probabilities of a student taking that module achieving that grade into 5% wide bins (0–5%, 5–10%, …,95–100%) and record the proportion of module results in that bin which obtained that grade. For a perfect prediction, we would expect the proportion of students that were given that grade to lie within the grade boundary of the bin (i.e. we would expect 10% of module results predicted to have a 10% probability of failure to actually fail). To measure how far away the predictions are from the correct results, our modified-RSS score is as follows:

$$\sum_{i=1}^{b} n_i(\overline{m}_i - p_i)^2$$

Where b is the number of bins, which with 5% sized bins is 20, n_i is the number of module instances in bin i, $\overline{m}_i$ is the midpoint of the bin i (e.g. 0–5% has a midpoint of 2.5%), and p_i is the proportion of module instances in bin i that achieved the grade analysed. The modification from the commonly used RSS score comes in the form of the weighting n_i, which prevents bins that have low counts from being too influential, and giving more significance to the more densely populated bins.

Without any information on VLE usage or other engagement, predictions on the following year's results would be based purely on the proportions of module results in each grade boundary in the previous year. This provides us a ‘base’ model where an example of our inference would be ‘if 10% of modules failed last year, then assign a 10% probability to all modules instances failing this year’. From this model, we can determine if VLE usage adds value to prediction, by comparing our modified RSS to a base RSS shown below:

$$n(p - q)^2$$

Where n is the total number of module instances, p is the proportion of module instances that resulted in the grade analysed in 2016/17 and q, the proportion of instances that resulted in the grade in 2015/16. If our modified RSS score is lower than this base RSS score, it suggests that the predictions using VLE are on average correct more often than without. In section 4.5, we refer to our models as ‘Prior’ and ‘VLE’ for the models without and with VLE usage included respectively.

4. Results

4.1. Differences in VLE usage across disciplines

A qualitative inspection of the different modules' usage of the VLE shows that there is variation in the ways VLEs are used by the

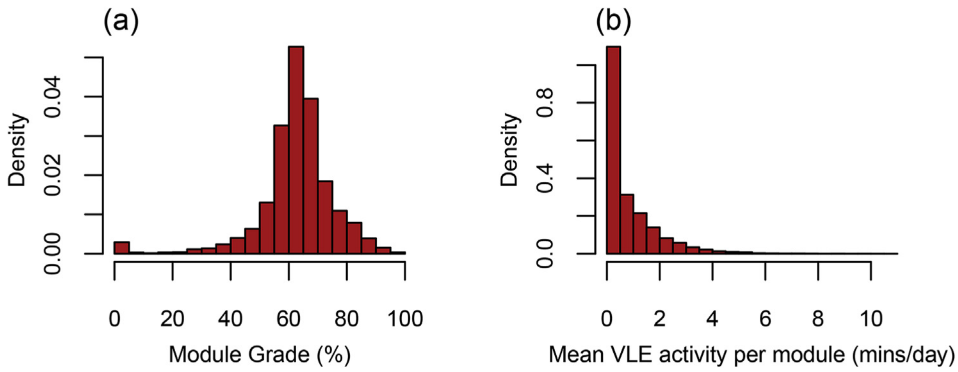


Fig. 1. The distributions of (a) module results and (b) mean term-time VLE activity per module for those in our study (see Table 1).

module instructors. This variation appears to be at a programme or College level. In other words, modules in the same College appear to similarly use the VLE. Specifically, we find that nearly all modules we looked at used the VLE to share the lecture slides with the students. Biological and Medical Sciences modules use the VLE for assessment purposes, by providing students with information about practical sessions (experiments, etc.) that make up the part of the students' final module grade, and also to carry out some summative assessments (e.g., multiple choice questions). We found the VLE pages for English modules were also concentrated around assessments, but in a more informative way. This includes providing information for assessments that the students would write away from the platform, as well as providing resources on how to write essays and use the library. VLE pages for Mathematics modules aside from lecture slides used the VLE to refer students to exercising resources (such as worksheets for the students to work through offline, as well as past exam papers). Politics modules used the VLE to refer students to extra reading resources, which were linked to each set of lecture slides. Economic modules had similar, although less, additional resources and worksheets similar to Mathematics.

4.2. *Relationships between VLE usage and module grade (whole cohort)*

We begin by looking across our whole study group, without taking into account differences between programmes. Fig. 1 shows the full module results (Fig. 1a) and mean VLE minutes of activity per day (Fig. 1b) for all 8239 student/module instances in our study. Module results are reasonably normally distributed (Fig. 1a), with a negative skew most likely due to students who failed to continue with the module after only getting part way through the term. VLE activity (Fig. 1b) appears to show a Pareto or Zipf's distribution often associated with levels of activity (e.g. Adamic and Huberman (2003)), though we do not quantify the fit to these distributions. It suggests that a lot of people have a small level of interaction and a small amount of people have a lot of interaction.

Plotting these against each other (Fig. 2) shows more clearly the distribution of VLE activity across students with different levels of grade performance. It appears that a majority of students interact very little with VLE (e.g. < 2 min per day) and still get good marks (e.g. > 60%). There are also few students who have high VLE activity that do not have a high grade. A chi-squared test that separates high and low VLE usage and high and low success (using the median values of 1.12 min and 64% respectively as separators) shows that this difference is significant ($p < 0.001$), with far fewer students falling in the high-VLE/low-grade quadrant than the null expectation. The overall Spearman's rank correlation between VLE usage and module result is 0.262 ($p < 0.001$). However due to the suggestion from the chi-squared test result that higher VLE usage is beneficial to getting a higher mark and is not related to low grades, we also split the students into high and low performers using the pass mark (40%) as a boundary and measure the Spearman's rank correlation coefficient separately. We find a stronger correlation between VLE usage and module results in students with grades below 40% ($r = 0.497$, $p < 0.001$) compared to students with grades above 40% ($r = 0.298$, $p < 0.001$). The stronger correlation for low performers could suggest that more interaction with VLE might improve results for struggling students (with the standard caveat that our results show only correlation, not causality, in this relationship).

4.3. *VLE usage and module grade within programme*

When modules are grouped by programme, we observe clear differences between programmes in the range and spread of student VLE usage associated with each programme (Fig. 3, Table 3). For example, there are students in BSc Biological Sciences, Mathematics, Economics and Medical Sciences who have mean daily VLE usage over 8 min, whereas there is only one person in each of BA English and Politics who have mean usage over 5 min. Furthermore, we find a much larger spread in overall module grades for the BSc programmes, which typically range from around 20 to over 90, compared to the BA programmes which show very few grades above 80 or below 40 (besides those students that get no mark at all, which we assume to be withdrawals or other exceptions). These qualitative observations are confirmed by inspection of the summary statistics given in Table 3, which show that mean grades are relatively similar for all programmes, whereas mean daily VLE usage shows greater variation. Variation in grades and VLE usage varies substantially between programmes.

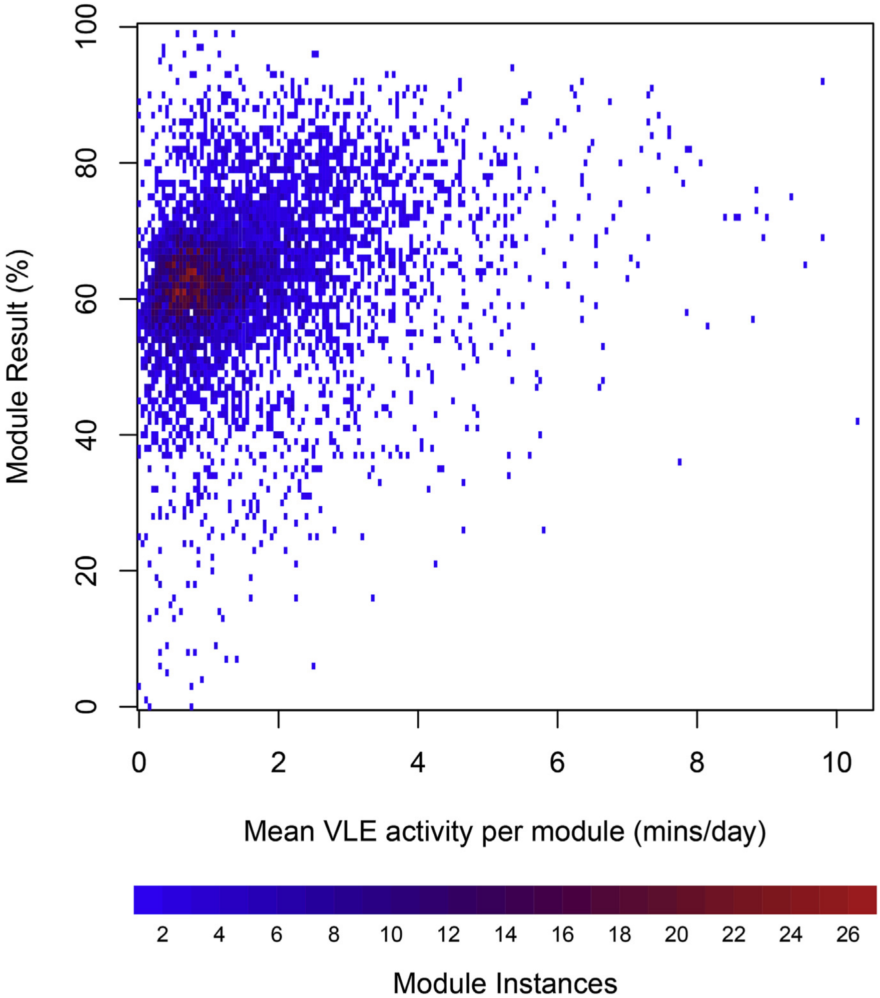


Fig. 2. Heatmap of the mean number of minutes of VLE activity per day for each module against the result from that module (%). Mean minutes are calculated over the term or terms that the module was taught across.

4.4. Individual variation in VLE usage (within module)

To illustrate how students might engage differently on VLE, we look at four typical students who take an example module (Ecology) in BSc Biological Sciences (Fig. 4). Within this module, which was chosen to give a large spread in activity/grades, we select a student from each quadrant: high-grade/low-activity, high-grade/high-activity, low-grade/low-activity, low-grade/high-activity. Looking at the time series of daily VLE usage by these four students, it appears that the two students who receive high marks have much more consistent activity patterns, whether it be low or high. The low-grade/low-activity student only logs onto VLE 12 times in the whole term. The low-grade/high-activity student has a much greater total usage, but their activity is still sporadic.

4.5. Learning design explains variation in strength of relationship between the VLE usage and grades

When we measure the strength of (Spearman's) correlation between VLE activity and grades for each individual module, we find a range of strengths in the relationship (Fig. 5). When we group the module-level correlations by programme, it is clear to see that there is a relationship between programme and correlation strength. The BSc programmes in Biological Sciences, Economics and Medical Sciences have relatively high correlations (mean correlation scores across modules of 0.33, 0.39 and 0.52 respectively) compared to the BA programmes in English and Politics (0.15 and 0.21 respectively).

The programme-level differences in correlation strength might be explained by variation in the styles of teaching and assessment that are commonly used in each discipline (Table 2). To determine whether the strength of module-level correlation between activity and grades was explained by differences in LD, we ran Spearman's correlation tests between the module-level aforementioned correlation strength (here treated as a variable), the mean and standard deviation of daily VLE usage, and the LD variables presented in Table 2.

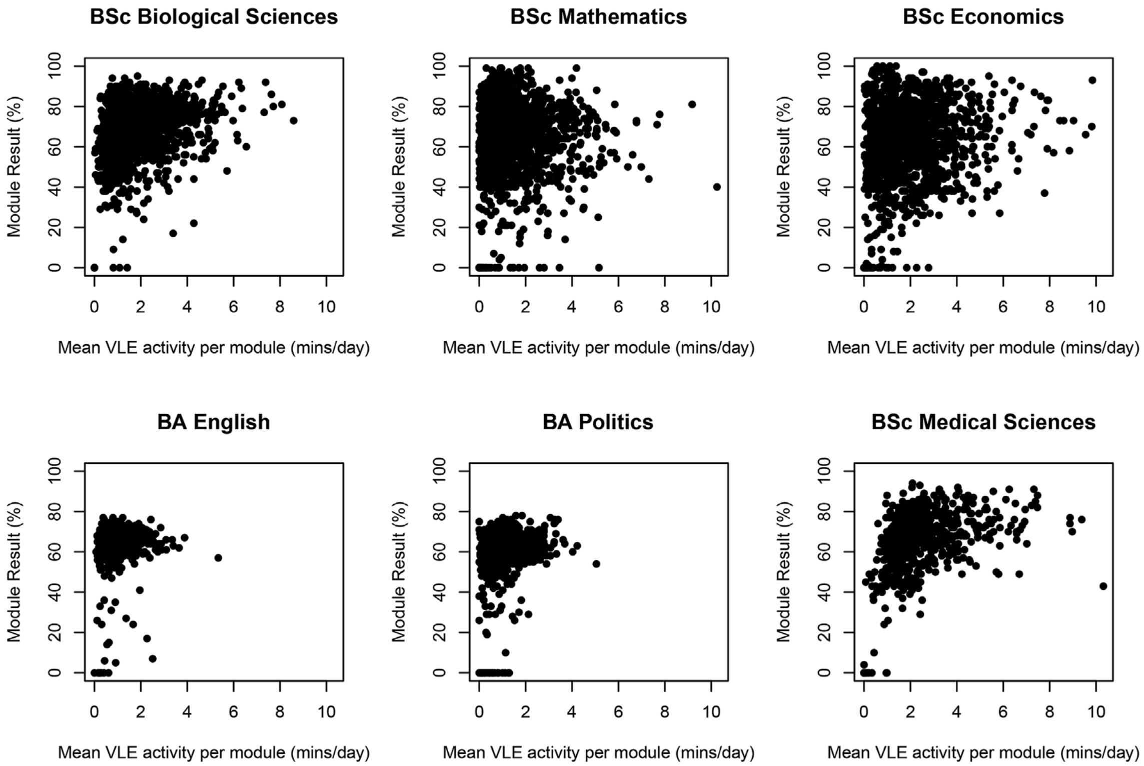


Fig. 3. Mean daily VLE activity per module vs module result (%) when modules are grouped by programme (see Table 1). As in Fig. 2, mean daily VLE usage is calculated across the term or terms the module was taught in.

Table 3
Mean and standard deviation (SD) for module grade (percentage) and VLE activity (mean daily minutes of usage per student) by programme.

Programme	Mean Grade (%)	SD Grade (%)	Mean VLE Activity (mins)	SD VLE Activity (mins)
BSc Biological Sciences	68.068	13.788	1.935	1.193
BA English	64.266	8.699	0.916	0.536
BSc Mathematics	63.252	18.415	1.377	1.219
BA Politics	60.366	9.729	0.969	0.616
BSc Economics	62.247	18.096	1.915	1.438
BSc Medical Sciences	65.415	15.922	2.544	1.551

The number of hours allocated to "Scheduled Teaching Activities" was found to positively correlate with the strength of correlation between VLE activity and module grade ($r = 0.336$, $p = 0.039$), and with the mean ($r = 0.538$, $p < 0.001$) and standard deviation ($r = 0.501$, $p < 0.001$) of VLE usage. Similarly, the number of hours allocated to "Guided Independent Study" (the complementary component of teaching time allocation to Scheduled Teaching Activities) was found to negatively correlate with the strength of correlation between VLE usage and module grade ($r = -0.336$, $p = 0.039$), and with the mean ($r = -0.538$, $p < 0.001$) and standard deviation ($r = -0.501$, $p < 0.001$) of VLE usage.

In addition, the weight of essays in the summative assessment for each module was significantly negatively correlated with the strength of correlation between VLE usage and grades ($r = -0.399$, $p = 0.013$), and with the mean ($r = -0.540$, $p < 0.001$) and standard deviation ($r = -0.609$, $p < 0.001$) of VLE usage. Thus, a higher weight of assessment by an essay means a weaker correlation between VLE and grades and a lower usage level.

Thus overall we find that LD may be an important explanatory factor for the observed differences between programmes/disciplines in terms of the strength of relationship between student engagement with the VLE and their module grades. Modules with more scheduled teaching activities (i.e. more contact time) and less weight on essay assessments, suggestive of scientific disciplines, show a stronger relationship between VLE usage and module grades.

4.6. Predictive power of VLE usage to determine grade

The differences in the grades predicted under varying VLE usage in each OLR model are shown in Fig. 6. Fig. 6 shows, for example, that Biological Sciences students who spend little time on VLE (i.e. near 0 min daily) have approximately a 7% chance of

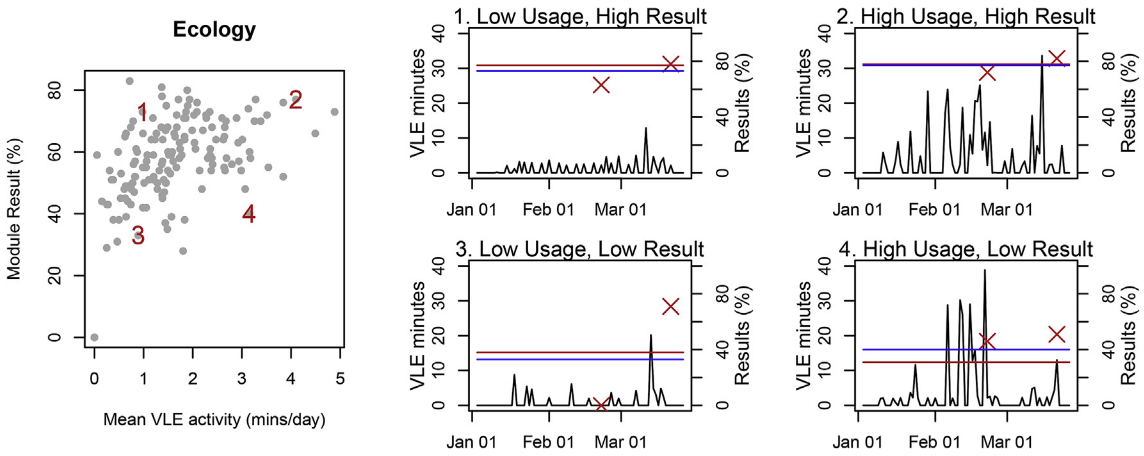


Fig. 4. Example student time series for a chosen BSc Biological Sciences module ('Ecology'). Students '1', '2', '3', '4', as noted in the scatter plot have their time series of activity shown respectively. VLE activity in minutes per day is shown as the black time series. Red crosses denote term-time coursework hand-in dates and the mark that was received (right hand axis). The red line refers to the end of year exam students take sometime after term (see Discussion) and the blue line denotes the overall mark the student received in the module. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

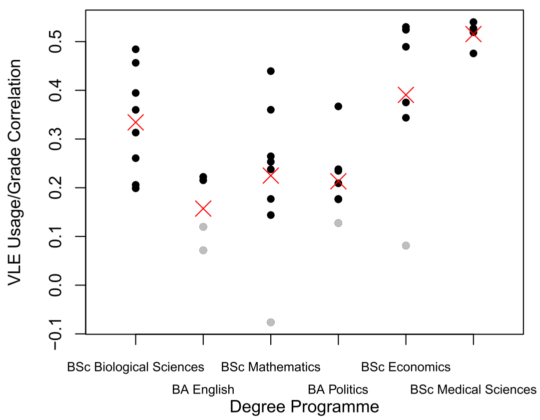


Fig. 5. Correlations between VLE activity for a specific module and the results in that module. Black dots represent the individual modules and crosses, their average across the programme they are part of. Correlations that are not significant ($p > 0.05$) are shown in grey, but are still included in the calculation of the mean.

failing their module, and relatively equal probabilities of getting any grade between a 3rd and a 1st. However, increased daily VLE usage is associated with decreased probabilities for both the 3rd/2:1 and 2:2 categories, and increased probability of getting a 1st. In contrast, for English there is a high probability of getting a 2:1 without the use of VLE (~75%), with increased VLE usage associated with decreased probability of a 2:1 and increased probability of a 1st. For Mathematics, where we see a generally weaker relationship between VLE usage and grade, we find that increased VLE usage does not change the probabilities for each grade category as much as for some other degree programmes. The model for Politics predicts the most likely outcomes are either a 3rd/2:2 or 2:1 with no VLE usage, but improve to a 2:1 or 1st with high VLE usage. The dependence on VLE for Economics students is similar to that of those in Mathematics. The strongest case for using VLE appears to be in Medical Sciences, where getting a 1st with no VLE usage is the second lowest probability outcome. However, with high VLE usage getting a 1st is by far the most likely outcome.

To determine how well our models could predict a student's grades for a given module, we use our models from the 2015/16 academic year to predict grades from the 2016/17 year. Taking Biological Sciences as an example to visualise what is detailed in the Methods section, for each grade category, we bin the probabilities of the module outcome being a given grades into 5% brackets (i.e. 0–5%, 5–10%, …, 95–100%) and in each bin determine what proportion of those were awarded the grade in question. We could envision this model being used in an operation context to monitor student performance over a teaching semester, with greater data volume available as input as time passes. Fig. 7 shows the how well the model predicts students' module results in this hypothetical context by using the mean daily VLE usage up to the end of week 1, week 5 and week 10 in the term. Results are presented by plotting the midpoint of the 5% bracket against proportion of students who got the predicted grade in each category. As such we would expect a good prediction to show points plotted on the diagonal dotted lines. The horizontal line refers to the proportion of students who got that grade overall last year, regardless of which 5% bracket they were predicted to be in. This represents the information that we

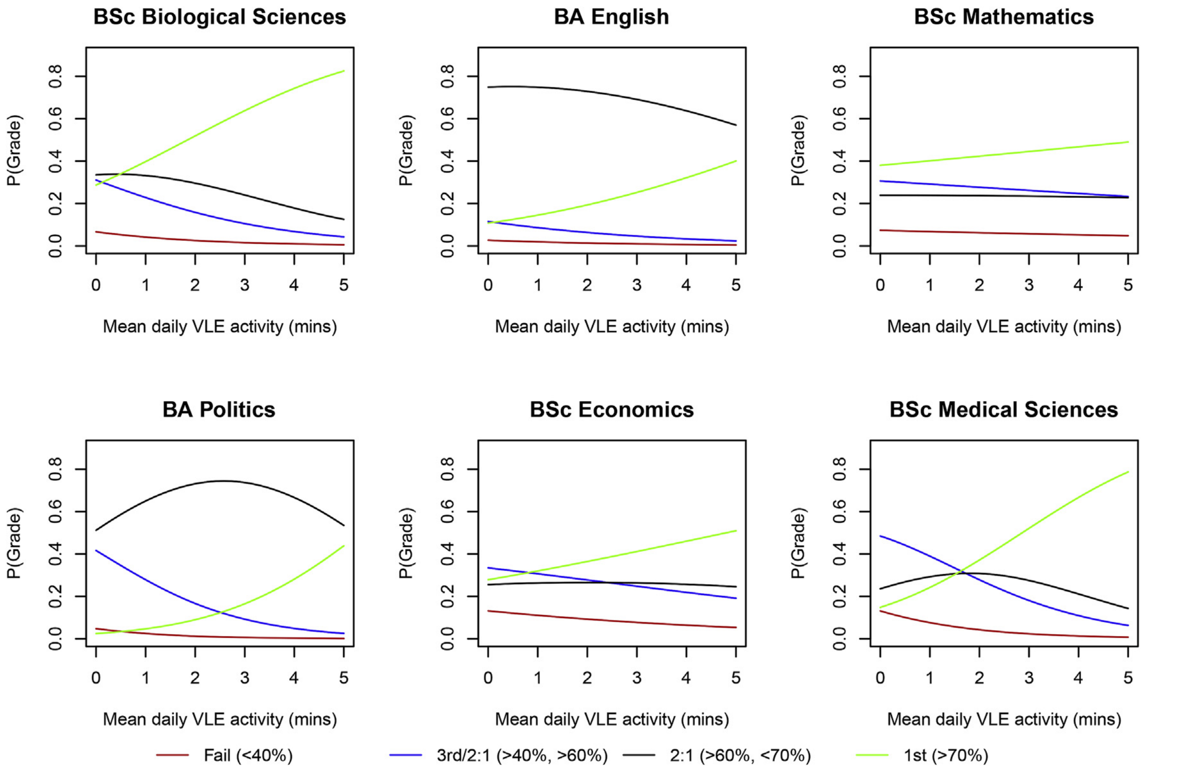


Fig. 6. Output from ordinal logistic regression (OLR) models on our 2015/16 dataset, separated by degree programme. Lines show the probabilities of being in each grade category (see main text) given the mean number of minutes the student spend on VLE per module.

would have without any information on VLE usage.

Fig. 7 shows that even after the first week, there is some predictive power of our model (shown by how the points begin to line up diagonally in the left column). However, with more input data available (e.g. after 5 weeks, middle column, and after 10 weeks, right column), we find the predictive power increases, with the points moving closer to the diagonal line. We also see that those points which contain more predictions are generally closer to the diagonal line and points off this are less populated. This is because bins with a relatively small number of module instances in them have a limited number of potential proportion of correct predictions in them (e.g. a bin with 5 module instances in can only have 0, 0.2, 0.4, 0.6, 0.8 or 1.0 as a proportion of correct predictions) and thus are more difficult to line up. Nonetheless, that our model predicts more densely populated bins well suggests it works for a large proportion of the population.

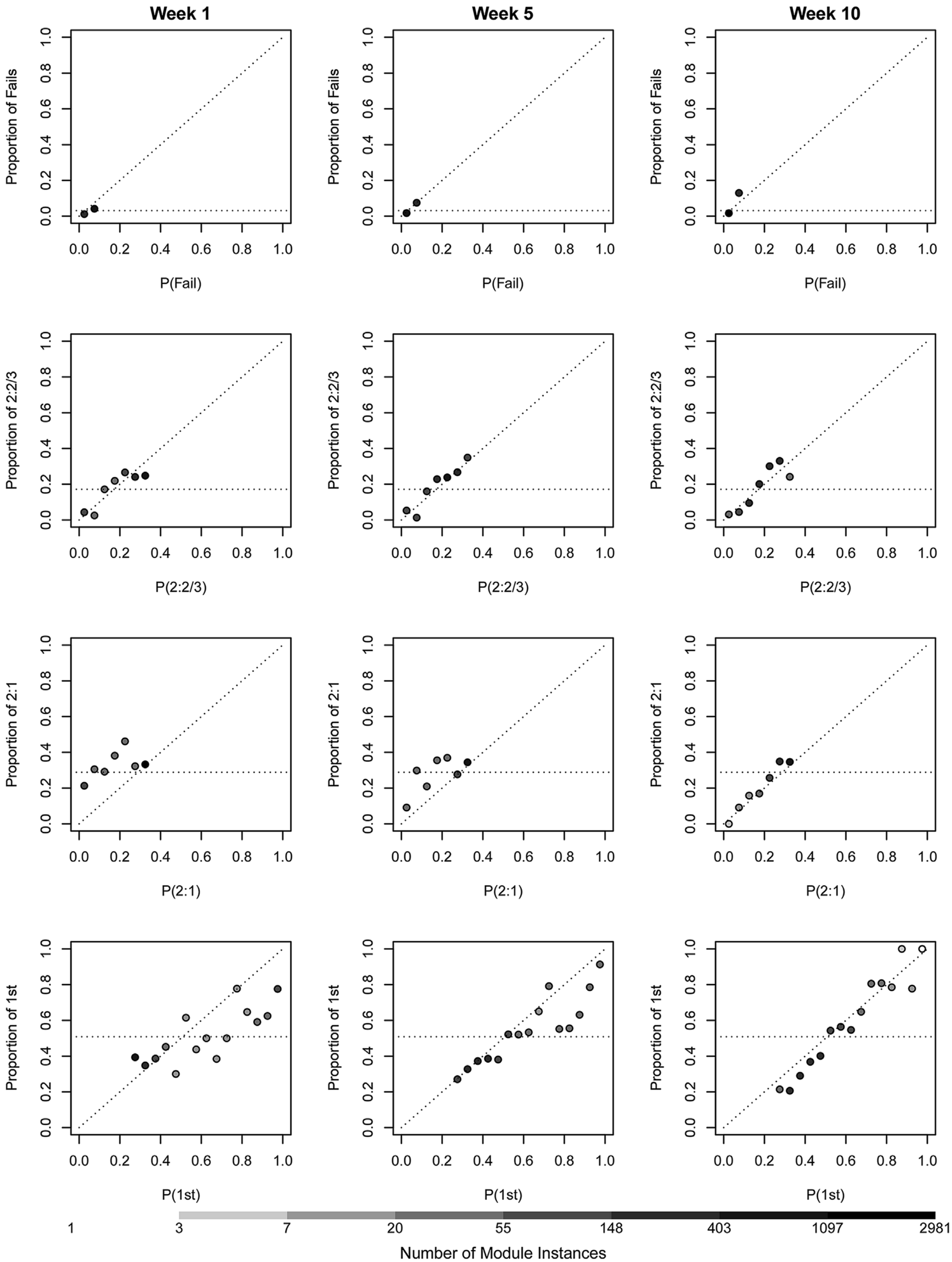
The base (Prior) and modified (VLE) RSS scores described in the methods are shown for VLE usage after 10 weeks for all 6 degree programmes in Table 4, with cells starred if adding VLE usage provides more information in predicting grades, and blue if not.

Table 4 shows that for the majority of cases, adding VLE usage does not provide better prediction (e.g. all of the RSS scores are lower by using last year's predictions for Medical Sciences). However in certain cases it could be a benefit. For example, having VLE usage data in Biological Sciences creates better predictions for 2:1 and 1st outcomes, and for all but failing in English. This matches our earlier findings that students who work on the VLE a lot get good marks, but not working on the VLE does not mean the student will not get good marks (and thus the lower grades are less predictable). Contrasting this, we find that a lack of VLE usage can be a good predictor of fails in Mathematics and Economics.

5. Discussion

Our study has looked at the relationship between student engagement (VLE usage) and learning outcomes (module grades) at a bricks-and-mortar university in the UK. We find a wide range of relationships between grades and use of the VLE, with variation for high-/low-performing students and between disciplines. This contrasts with online-only and distance-learning contexts, for which VLE usage has been shown to be more strongly and consistently linked to achievement, such as Wolff, Zdrahal, Herrmannova, Kuzilek, and Hlosta (2014) and Kuzilek et al. (2015). In these cases, the strong relationships are found in an environment where the VLE is the main point of interaction between students and instructors. In a bricks-and-mortar university (such as the UoE examined in this study) this is not the case, with the majority of teaching conducted face-to-face through lectures and seminars, and self-study involving 'offline' interactions such as visiting the library or working with peers.

Across all programmes, we find that high VLE usage in a module is associated with high grades, but it does not follow that low VLE usage is necessarily associated with low grades. In fact, it appears that the majority of students do not engage with VLE very



(caption on next page)

Fig. 7. Prediction of BSc Biological Sciences module results based on VLE usage from 2016/17 academic year, using the OLR model coefficients from the 2015/16 academic year. Predictions are grouped into 5% breaks (0–5%, 5–10%, …) and plotted against the proportion of module results that resulted in the correct grade category. Points are coloured depending on how populated each break is. Perfect prediction would line up on the dotted diagonal line. The dotted horizontal lines show the overall proportion of modules that had that grade category outcome, and as such show the prediction that would be made with no prior knowledge of VLE usage.

Table 4
Base/Prior and modified/VLE RSS scores as described in the Methods. Pairs of cells are starred if the VLE usage provides better prediction.

Degree	Fail		3rd/2:2		2:1		1st	
	Prior	VLE	Prior	VLE	Prior	VLE	Prior	VLE
BSc Biological Sciences	0.01	0.71	2.88	3.63	2.03*	1.84*	10.44*	7.29*
BA English	< 0.01	0.03	1.19*	0.51*	1.47*	1.39*	5.37*	4.83*
BSc Mathematics	4.00*	2.77*	0.05	0.71	0.10	0.11	2.10	7.02
BA Politics	1.51	1.71	2.35	7.19	11.17	29.35	13.30	15.69
BSc Economics	4.16*	3.22*	0.49	3.77	0.01	1.17	6.87	13.14
BSc Medical Sciences	0.20	8.69	0.43	5.51	1.62	3.75	1.13	4.63

much and still get good grades. This suggests that these students are engaging with learning in ways that do not involve the VLE and which are not captured by our focus here on VLE usage. This finding is counter to the original expectation that VLEs and Learning Management Systems will transform education (Black, Beck, Dawson, Jinks, & DiPietro, 2007; Coates, James, & Baldwin, 2005). Also, it may suggest that other individual factors, such as prior knowledge, account for the outcome of learning (Kennedy, Coffrin, Barba, & Corrin, 2015).

We can anecdotally link the way VLE is used by module lecturers and how much this is related to students' results in those modules. It is clear that Biological and Medical Sciences students need to log on each week to see what their practical sessions involve and complete tests that go towards their final grade, thus their VLE usage is directly linked to their assessment and outcome as a result. However, English students are more likely to get the assessment information they need and complete it offline, with less assessments to look up on VLE compared to the numerous practical assessments in Biological and Medical Sciences. Thus, it seems likely that usage of VLE will be less related to academic outcomes in English. Mathematics students are likely to get worksheets that require pen and paper, with only a few for each module, so there is little need for them to spend a lot of time on VLE. The same is true of Economics. The extra information available to students on the VLE Politics pages appeared to be a very useful resource and thus students who read around the subject more are likely to get higher grades, which might indicate that reading is strongly related to their outcome. Practically, tutors could encourage students who show early signs of poor performance to increase their use of the VLE.

We also find substantial variation between disciplines/programmes in the strength of relationship between VLE usage and grades, with stronger relationships found for BSc Biological and BSc Medical Sciences than for BA English and BA Politics. Analysis of the LDs used in these programmes/disciplines helps to explain this difference. Our results show that the strength of relationship between VLE usage and grades depends positively on the amount of scheduled teaching activities (contact time) and negatively on the time assigned to independent study. Thus it seems that (at the UoE) the VLE is a better predictor of outcomes in LDs which are more instruction-based by nature ('sage on the stage', face-to-face learning), and a worse predictor for LDs which are based on self-study. In addition, we have shown that weight assigned to an essay writing as a summative assessment tool, is negatively correlated with the predictive power of the VLE usage, suggesting that the assessment's design is another LD characteristic that highly affects the usage levels and prediction ability of the VLE. We also find a larger spread in VLE usage in programmes that have a LD tailored more to instruction-based learning, suggesting that (perhaps counterintuitively) there is less benefit in accessing VLE content in those programmes where the LD encourages self-guided learning. This is most likely because the self-guided learning is mainly done "offline" by reading from books and other sources suggested by the reading list uploaded to the VLE by the lecturer.

Regarding our predictive OLR model, it is important to note that it does not predict the most likely outcome, and that its real strength is in assigning probabilities for all grade categories. For example, we note that for predicting fails, there is only a maximum probability of 10% assigned (Fig. 7). However, although this is not a likely outcome, 10% of these module outcomes fail. This could be useful to tutors who could see a student is more at-risk of failing that module than another student who might be predicted a much lower probability. We also note here that the relationships shown in Fig. 6 do not necessarily reflect the correlations observed in Fig. 5. This is in part because the mean number of VLE minutes per college is much lower than we have used as a range in our model (Table 3). Also the non-linearity in the relationships found in Fig. 3 is picked up by these models, but not so much by the correlation values, although we reiterate we are using Spearman's correlation coefficient which works best for these non-linear relationships. Nonetheless, our models show the general behaviour we observed in Fig. 3, that students who do not work on VLE do not necessarily do badly, but those who do use VLE generally do better. One caveat to our results is a temporal separation for some modules between the term-time learning period in which we measure VLE usage and the end-of-module examinations that account for a significant proportion of the students' final module grades. At the UoE, most exams are taken in May/June, irrespective of whether the module was taught in the Autumn term or the Spring term. Thus our measurement of VLE usage captures only the period during which material was taught and learned, and not the revision period during which much of the assessment occurred. In addition, some modules/programmes (e.g. for BA English) place less emphasis on end of year exams than others (e.g. BSc Biological Sciences), which creates variation in the temporal separation between learning and assessment, adding another possible confounding factor. We also find a smaller range of module results in the BA programmes, which could make them harder to correlate with engagement, compared to the larger spread in BSc programmes.

Another limitation is that while we have tried here to capture a representative sample of students and modules from across the disciplinary spectrum, we still only measure engagement and performance on a small fraction of the ∼3000 modules available at the UoE. The sampled modules were chosen such that there was a large number of students taking them, leading us to focus on first-year compulsory modules from programmes with large enrolment. Relationships may differ in other modules, or in other years of study, or in postgraduate programmes. Relationships in general will be a lot harder to determine for modules with smaller numbers of students enrolled.

Finally, students all learn in different ways (e.g. Keirsey and Bates (1984)) and as such we might expect different dependencies on VLE usage for different students in the same module because of their different approaches to study. While our analysis of individual student VLE usage patterns here has uncovered an interesting suggestive finding that high-grade students have more consistent usage patterns (irrespective of their total usage volume), these results are anecdotal and we have not fully explored individual-level patterns.

The finding here that VLE usage alone does not appear to be a complete predictor of student performance strongly suggests that VLE usage does not capture all aspects of student engagement with learning and highlights the difficulty in diagnosing engagement in the diverse learning environment of a bricks-and-mortar university. A better measurement approach would need to include a wider variety of sources, combining not just VLE usage but also (e.g.) library usage, attendance at lectures and other scheduled activities, usage of other digital systems, etc. The predictive power of other metrics is not immediately clear (Kent, Boulton, & Williams, 2017). Furthermore, our results suggest that different forms of engagement may be more or less predictive in different contexts (e.g. different disciplines, programmes, modules, or different parts of the cohort within a single module). Thus different weightings may need to be associated with each form of engagement and these may need to be re-calculated for each module in turn and possibly for different instances of the same module.

6. Conclusion

Many universities are using VLE activity as a primary metric and proxy for engagement with learning, with various studies of how such online engagement is related to student performance. The relationship between engagement and learning outcomes is complex, especially in bricks-and-mortar universities where teaching and learning take many forms, both online and offline. It is important for higher education managers and policy makers to understand exactly how and under which conditions this relationship manifests itself. Here we present evidence showing that student grades are related to VLE usage, but that the relationship is not straightforward. We have shown that at a 'bricks-and-mortar' university, where student interactions with learning are varied in nature and include both face-to-face and online activities, that levels of engagement with the VLE vary depending on the course/discipline the student is studying. Furthermore, we find that the importance of VLE usage for prediction of a student's grade outcome differs between courses/disciplines. While this appears to be partially due to the learning designs for different courses, we also find evidence for differences in how departments utilise VLE in their learning activities. Being aware of these disciplinary differences in utilisation, as well as the differences in importance of VLE usage for student outcomes, should be of great benefit to universities in designing effective teaching practices. This is particularly true of those which offer predominantly face-to-face learning, and where a VLE is an auxiliary teaching platform, rather than in distance-learning contexts where a VLE is the main platform for student learning.

Acknowledgments

This research was supported by the Effective Learning Analytics project at the University of Exeter. This project aims to make effective use of data to help students reach their full academic potential while studying at University of Exeter.

References

Adamic, L. A., & Huberman, B. A. (2003). Zipf's law and the internet. *Glottometrics, 3*, 143–150.
Adeyinka, T., & Abdulmumin, I. (2011). Pattern of undergraduate's participation in the online discussion forum at the University of Ilorin, Nigeria. *Journal of Information Technology Management, 22*(3), 59–76.
Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M.Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior, 31*, 542–550.
Ahlfeldt, S., Mehta, S., & Sellnow, T. (2005). Measurement and analysis of student engagement in university classes where varying levels of PBL methods of instruction are in use. *Higher Education Research & Development, 24*(1), 5–20.
Akyol, Z., & Garrison, D. R. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Educational Technology, 42*(2), 233–250.
Andresen, M. A. (2009). Asynchronous discussion forums: Success factors, outcomes, assessments, and limitations. *Educational Technology & Society, 12*(1), 249–257.
Asterhan, C. S. C., & Hever, R. (2015). Learning from reading argumentive group discussions in Facebook: Rhetoric style matters (again). *Computers in Human Behavior, 53*, 570–576.
Axelson, R. D., & Flick, A. (2010). Defining student engagement. *Change: The Magazine of Higher Learning, 43*(1), 38–43.
Black, E. W., Beck, D., Dawson, K., Jinks, S., & DiPietro, M. (2007). The other side of the LMS: Considering implementation and use in the adoption of an LMS in online and blended learning environments. *TechTrends, 51*(2), 35–39.
Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student Learning: Testing the linkages. *Research in Higher Education, 47*(1), 1–32.
Casuso-Holgado, M. J., Cuesta-Vargas, A. I., Moreno-Morales, N., Labajos-Manzanares, M. T., Barón-López, F. J., & Vega-Cuesta, M. (2013). The association between academic engagement and achievement in health sciences students. *BMC Medical Education, 13*(1), 33.

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education, 96*, 42–54.
Coates, H., James, R., & Baldwin, G. (2005). A critical examination of the effects of learning management systems on university teaching and learning. *Tertiary Education and Management, 11*(1), 19–36.
DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing “Course”:Reconceptualizing educational variables for massive open online courses. *Educational Researcher, 43*(2), 74–84.
Ferguson, R., Clow, D., Beale, R., Cooper, A. J., Morris, N., Bayne, S., et al. (2015). *Moving through MOOCS: Pedagogy, learning design and patterns of engagement.* Cham: Springer International Publishing70–84.
Gunuc, S. (2014). The relationships between student engagement and their academic achievement. *International Journal on New Trends in Education and Their Implications, 5*(4), 216–231.
Harris, R. A., & Nikitenko, G. O. (2014). Comparing online with brick and mortar course learning outcomes:An analysis of quantitative methods curriculum in public administration. *Teaching Public Administration, 32*(1), 95–107.
Keirsey, D., & Bates, M. (1984). *Please understand Me: Character and temperament types.* Prometheus Nemesis Book Company.
Kennedy, G., Coffrin, C., Barba, P. d., & Corrin, L. (2015). Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance. *Proceedings of the fifth International conference on learning Analytics and knowledge* (pp. 136–140). Poughkeepsie, New York: ACM.
Kent, C., Boulton, C. A., & Williams, H. H. T. (2017). Towards measurement of the relationship between student engagement and learning outcomes at a bricks-and-mortar university. *Sixth multimodal learning Analytics (MMLA) workshop and the second cross-LAK workshop co-located with 7th International learning Analytics and knowledge conference (LAK 2017). Vancouver, Canada*.
Koper, R. (2006). Current research in learning design. *Educational Technology & Society, 9*(1), 13–22.
Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., & Wolff, A. (2015). OU Analyse: Analysing at-risk students at the open university. *Learning Analytics review, LAK, 15*(1), 1–16.
Lehmann, J., Lalmas, M., Yom-Tov, E., & Dupret, G. (2012). *Models of user engagement.* Berlin, Heidelberg: Springer Berlin Heidelberg164–175.
McConnell, D. (2000). *Implementing computer supported cooperative learning.* Psychology Press.
Millar, R. (2013). Improving science Education: Why assessment matters. In D. Corrigan, R. Gunstone, & A. Jones (Eds.). *Valuing assessment in science Education: Pedagogy, curriculum, policy*. Springer.
Nguyen, Q., Rienties, B., & Toetenel, L. (2017). Unravelling the dynamics of instructional practice: A longitudinal study on learning design and VLE activities. *Proceedings of the seventh International learning Analytics & knowledge conference* (pp. 168–177). Vancouver, British Columbia, Canada: ACM.
Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, R., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior, 76*, 703–714.
Picciano, A. G. (2002). Beyond student Perceptions: Issues of interaction, presence, and performance in an online course. *Journal of Asynchronous Learning Network, 6*(1), 21–40.
Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior, 60*, 333–341.
Shavelson, R. J., & Huang, L. (2003). Responding responsibly. *Change: The Magazine of Higher Learning, 35*(1), 10–19.
Song, L., & McNary, S. W. (2011). Understanding student's online interaction: Analysis of discussion board postings. *Journal of Interactive Online Learning, 10*(1), 1–14.
Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157–167.
Wilson, K., & Fowler, J. (2005). Assessing the impact of learning environments on students' approaches to learning: Comparing conventional and action learning designs. *Assessment & Evaluation in Higher Education, 30*(1), 87–101.
Wolff, A., Zdrahal, Z., Herrmannova, D., Kuzilek, J., & Hlosta, M. (2014). Developing predictive models for early detection of at-risk students on distance learning modules. *Machine learning and learning Analytics workshop at the 4th International conference on learning Analytics and knowledge (LAK14)* Indianapolis, Indiana, USA.
Zhu, E. (2006). Interaction and cognitive engagement: An analysis of four asynchronous online discussions. *Instructional Science, 34*(6), 451.